Melodic Pattern Recognition and Similarity Modelling: A Systematic Survey in Music Computing

Makarand Velankar[1], Parag Kulkarni[2]

[1]Information Technology Department, Mksss's Cummins College of Engineering for women, SPPU, Pune, India
[2]Iknowlation Research labs, Pune, India

E-mail: [1]makarand.velankar@cumminscollege.in, [2]parag.acad@gmail.com

Abstract

Musical pattern identification is crucial for various classification and retrieval applications in computational musicology. Feature learning is the basic task, and features act as a basis for the Pattern Recognition (PR). Selecting an appropriate approach is vital to the accuracy of the retrieval algorithms. This research gives a comprehensive review of approaches used for PR and similarity modelling. It systematically analyses various approaches for melodic feature identification and comparatively evaluates the work done in the literature in terms of software tools used, melodic pattern representations, and matching. The study discusses the benefits and limitations of various approaches along with the challenges to be addressed for melodic PR. Results show a wide variety of approaches for different music genres and applications. Further, analysis shows that statistical and symbolic approaches were being used predominantly, and deep learning approaches are gaining popularity in recent times.

Keywords: Pattern recognition, music analysis. computational musicology, melodic similarity

1. Introduction

Music is ubiquitous, and primarily used for entertainment to express emotions and themes. Today, music lovers or creators use the Internet to generate and upload music files in audio or video format. With smartphones, social media and the Internet, this music data is growing at an alarming rate. Moreover, computer-generated automated music with artificial intelligence also contributes to the rapid growth of music data (Dhariwal et al., 2020). This enormous growth of music data poses challenges for various music apps and websites to stay

relevant and update to cater user needs. Music retrieval and recommendation uses music pattern matching techniques to retrieve similar music. Patterns are dominant in music, with representative examples such as melodic (Deutsch and Diana, 1979) or rhythmic patterns (Shmulevich et al., 2001). Melodic patterns refer to ordered sequences represented by a discrete notation as symbolic representation (Cambouropoulos et al., 2001). Statistical representation is the time series of feature vectors, where the vector represents acoustic features calculated at regular time intervals (Klapuri, 2010). The work presented focuses on melodic pattern recognition, representation and modelling in music. Melodic pattern recognition has wide applications in music domain such as automatic transcription, content-based retrieval, copyright issues and claims by musicians etc.

World music can be classified as polyphonic, monophonic and homophonic on the broad lines considering melodic patterns (Salamon et al., 2015). Polyphonic music has more than one melody line being played simultaneously, whereas monophonic music has a single melody. Traditional nursery rhymes like "Twinkle, twinkle little star" is an example of monophonic music, whereas Mozart's or Bach's Fugue western music is an example of polyphonic music. Musical patterns used in folk or film music are of homogeneous type, predominantly (Sadie et al., 2001). A melody line is generally performed by the leading performer with the accompaniment of others in heterophonic music. Improvisations in melodic patterns during the performance are visible across music genres such as jazz music (Norgaard et al., 2022) or classical music in India (Chaudhuri, 2021).

The research in the Pattern Recognition (PR) domain typically uses a two-step approach in various applications (Bishop et al., 2006). The first step involves extracting features of the data. The second stage is feeding these features to the machine learning classifier as shown in Fig.1. Features used should maximize the distance between the means and minimize the variance within the class classification. Thus, the performance of classifiers depends on appropriate features, which is the crux of classification. A pattern is a physical object or abstract notion represented by a set of descriptors or features.

Musical pattern discovery algorithms (Ren et al., 2017) are widely researched topics specifically for western music. Conventional systems use feature extraction, selection, and, occasionally, feature engineering for music pattern analysis. A musical pattern is represented typically using a vector of feature values. Feature normalization is performed to provide equal importance to all features. A good model means using discriminative attributes and reducing the computational burden of classification. Considering the similarity and

adaptability of different machine learning technologies, some researchers believe that neural network-based classification methods are part of statistical methods. However, the neural network-based deep learning approach differs significantly during feature construction, where features are preserved in neuron weights and states. Recent advancements in computing have shown great potential in different musical tasks such as retrieval, analysis, recommendation, synthesis, and composition.
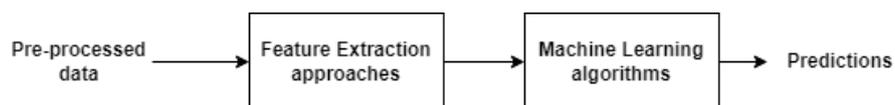
**Figure 1.** Typical process of pattern recognition

Traditionally the melodic patterns are represented using symbolic notations used by different music traditions and typical pattern recognition and matching approaches such as statistical and symbolic approaches are popular among research community. With the introduction of neural network based pattern recognition, especially deep learning based approaches have shown excellent results for images and speech domain. Musical patterns are complex and different techniques are being explored by researchers, including the conventional, novel and hybrid approaches.

The systematic survey of melodic PR in music computing is organized as follows. Section 2 covers the related work with different approaches used for PR and feature extraction along with software tools available. Section 3 deals with the discussion about the merits and limitations of the methods, and also records the challenges. Finally, section 4 presents the conclusion.

2. Related Works

Traditionally, different PR paradigms for music pattern analysis include statistical and structural approaches for a specific predefined task. The general classification of melodic PR approaches is as shown in fig. 2. A statistical approach (Jain et al., 2000) based on probabilistic models with efficient use of machine learning algorithms for different applications in music information retrieval is common among researchers. The structural or symbolic approach (Albus et al., 2012) based on a formal grammar (Roads and Paul, 1979) helps model melodic or rhythmic structures in the music. The use of neural networks for PR now extends to deep neural networks (Liu et al., 2017) for efficient prediction. The

boundaries between different paradigms are fuzzy and fading. Combined approaches are also gaining popularity as they share the same goals (Oramas et al., 2018).
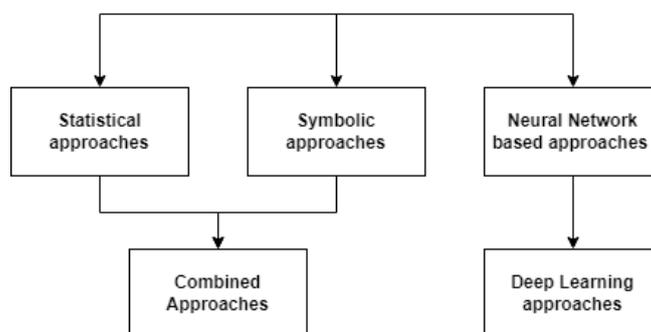
Figure 2. Melodic pattern recognition approaches

2.1 Statistical approach

As the name suggests, the statistical approach attempts to extract numerical values from the data as a source for classification. This technique has gained more acceptance and popularity in the research community due to different machine learning and classification algorithms (Kotsiantis et al., 2007) applied to the numerical data extracted from the digital objects under study.

Self-repetitive pattern identification is a topic of interest from music summary or content-based Music Information Retrieval (MIR). A comparative study of Chroma features, Constant Q transforms features, and MFCC features was performed. Results were compared with the ground truth obtained from human expert annotation for identifying repetitive patterns (Lu et al., 2004). The ground truth used in most of the systems is input from human experts. It is not easy to get the ground truth for large datasets with duration in hours for audio files. The challenge is to generate and evaluate ground truth for massive musical data, which can evaluate different machine learning algorithms. The vector space model was used for melodic pattern extraction of the raga in Indian art music, and the results were assessed for diverse classification strategies (Gulati et al., 2018). Melody track detection using statistical PR (Rizo et al., 2006) (Ozcan et al., 2005) was explored for MIDI music data. The statistical PR approach using vector representation is suitable for machine learning algorithms. They typically require input data in feature values. Modelling appropriate features of the music, which are interpreted as structures, is the challenge in music computing. The structural approach thus becomes necessary for melodic or rhythmic patterns, perceived on a timeline as a sequence.

2.2 Symbolic approach

Structural or symbolic PR is generally applied for melodic or rhythmic PR. In melodic patterns, the note sequence pattern is usually represented as an ordered list of notes with string type data structure. A monophonic musical pattern is described by using notes (Conklin, 2010) as the fundamental unit for the representation. The hierarchical tree structure is used to represent the pattern (Rohrmeier et al., 2018) at different granularity levels. The directed graph structure (Khulusi et al., 2020) is another representation of the notes transition in melody.

Induction and matching of sequential melodic patterns pose several computational challenges, but are helpful for musically interesting retrieval tasks (Benetos et al., 2013). Time series symbolic pattern representation of music is a challenge, as the data are multidimensional and real-valued, with patterns rarely repeating precisely. Data margins and fuzziness are used considering the perception of patterns. Pattern segmentation for processing can be done using perceptually grouped melodic phrases. Pitch transcription, rhythmic meter, or tempo information may not be the only helpful information for structural pattern analysis. The use of timbre information with string matching techniques can be more beneficial for polyphonic music (Aucouturier et al., 2002). The music structural pattern representation plays a significant role in automatic conversion from a sheet printed music to the performance aspect (interpretation) or vice versa (transcription). Evaluation of accuracies related to optical music recognition systems used for structural musical interpretation is a challenge (Bainbridge and Tim, 2001).

A graphical structure can represent music scores. The terminal nodes directly describe the content of the music, the internal nodes represent its incremental summary, and the arc represents its relationship. The similarity between two melodies can be calculated by analyzing the structure of the graph and finding the shortest path between corresponding nodes in the graph (Orio et al., 2009). Due to effective modelling of sequential patterns for the time-variant music and human perception of musical patterns, a structural approach is suitable for melodic patterns. Therefore, more robust music knowledge representation systems are possible with this paradigm. Combined approaches using statistical and symbolic approach are also gaining grounds as they compliment each other well to provide melodic info suitable for different applications (Pearce et al., 2018). Human brains process the musical patterns for different interpretations, and a neural network-based approach attempts to model the same.

2.3 Neural network-based approach

The artificial Neural Network (NN) is a computing model composed of interconnections of artificial neuron units to simulate the human brain. Different problems in music have been successfully addressed with the help of different variants of NN such as genre classification (Pelchat and Graig, 2020), teaching (Wang et al., 2021), and rhythm recognition (Chen, 2022). The use of dynamic programming and recurrent NN with hidden states or memory units for chord recognition is found better than the hidden Markov model (Boulanger et al., 2013). Convolutional NN trained on a Mel-scaled spectrogram showed promising results for automatic boundary detection for musical structures (Ullrich et al., 2014). Research indicates that NN can be trained to identify statistical inconsistencies across audio features to predict valence/arousal values for emotion classification (Vempala and Frank, 2012). With the popularity of deep learning, data-driven methods based on deep neural networks, are gaining more and more attention in the research of melody extraction.

Deep learning approach has produced excellent results in image recognition (Jagtap and Bhosale, 2021) and speech recognition (Subramanian et al., 2022). Deep learning for melody extraction is being explored recently by researchers. Lu et al.1, in 2018, investigated deep learning for music symbolic data and they further explored vocal melody extraction using patch-based CNN (Lu et al.2, 2018). CNN based two staged multi-resolution end-to-end model was proposed (Chen et al., 2019) for singing melody. A detailed review of melody extraction from polyphonic music by deep learning approaches (Reddy et al., 2022) is presented, and identified hindrances in the progress. Non availability of large datasets inclusive of different genres in the public domain for training and the need for standardizing the target pitch classes for universal melody extraction are major issues identified. After comparing the results, it was found out that there is much scope for improvement with new advanced architectures, input data preprocessing and training approaches.

2.4 Feature learning approaches

Feature learning plays a crucial role in machine learning prediction accuracies. Features are extracted using available tools or software libraries, or researchers develop new tools. The data needs to be in a digitized format to extract features. Unfortunately, not all Indian musical information is available in digital format. The data such as music notations or the lyrics are available as images or textbooks. Music digital objects are music info in digital formats such as audio files, text files for lyrics or notations or comments, videos, etc. The primary data source used is audio files in the majority of cases. Available tools and feature

selection approaches are studied and experimented with for audio data. Features play a crucial role in the successful implementation of machine learning algorithms. Considering the different dimensions of music and their interdependence with the various genres involved, modelling the characteristics of music is a challenging task. Therefore, a variety of features are used to train the machine for different music analytics tasks (Thickstun et al., 2016). With the growth of music data on the Internet, music data analysis has received much attention. It involves streaming, purchasing, uploading, and online music commenting specifically for popular music (Lee et al., 2018). The automatic recognition of the instrument or artist, style, genre, emotion, melody becomes necessary. Manual labor is time-consuming for music experts, and it is nearly impossible for humans to perform these tasks for a large amount of growing music data (Panteli et al., 2018).

Researchers proposed different melodic features and frameworks for various applications. Representative examples provide the utility and importance of melodic PR. For example, during the feature learning for the world music similarity task (Panteli et al., 2016), out of a total of 2170 recordings, 70 recordings were chosen at random from each of the 31 countries for feature extraction. Features used were onset patterns with the scale transform for rhythm, average chromatograms for harmony, and Mel Frequency Cepstrum Coefficients (MFCC) for timbre content description. In addition, melodic aspects were captured via pitch bi-histograms which denote counts of transitions of pitch classes.

Table 1. Feature approaches for Melodic Pattern Recognition

Task/ Application	Features used for experimentation	Dataset	Attainment
Music Similarity (Panteli et al., 2016)	Counts of transition of pitch classes in pitch bi-histograms	World Music	557 outliers out of 2169
Feature discovery (Jonas et al., 2017)	Viewpoint features: pitch, Interval, Contour and anchor features like a tonic key	Monophonic Music	322 useful features from 5851 features
Pattern similarity (Gonzalo, 2001)	Symbolic representation of melody and string matching	Folk Melodies	80% songs with pairwise distance > 0.7
Melodic Similarity (Tuomas et al., 2001)	Statistical measures as distribution of tones, intervals, duration and transitions	Western Music	predictive power of 52 %
Structural Analysis (Bee et al., 2006)	Pitch-Class Distribution (PCD) Pitch Class Profile (PCP) and Harmonic Pitch Class Profile (HPCP)	Songs The Beatles	Average accuracy of 86 %

Automatic Accompaniment (Xiang, 2009)	Pitch-Class Distribution (PCD)	Chinese Music	F measure 67.3%
MIREX Task (Justin, 2008)	Chroma based features with Levenshtein distance measure	Western Music	Extraction rate of 88.4 %
Cover song Identification (Samuel and Shrikanth, 2008), (Joan et al., 2008)	Chroma based features with Euclidean distance measure	Western Music	Accuracy of 79 % with improvement of 5.8 %
Melody Extraction (Klaus et al., 2019)	Minor to moderate n-grams 2 Gram, 3 Gram and 4 Gram	Jazz Music	Accuracy beyond 90 %
Melodic Classification (Darrell, 2009)	Tri-grams with best classification accuracy of 75% with 10-fold cross validation	Folk Music	Accuracy of 77 % for small dataset
Raga Identification (Rao et al., 2014)	PCD, N-gram, pakad as sequence of notes	Indian Classical music	Hit rate of 99% for HCM samples

Table 1 summarizes the different approaches used for melodic feature extraction by various researchers. Some important observations and conclusions drawn from them provided necessary directions. The performance of HPCP was observed slightly higher compared to others for repeated patterns (Ong et al., 2006). The chroma based approach was used for melody extraction during the MIREX (Music Information Retrieval Evaluation eXchange) task (Justin, 2008). During the experimentation with different window sizes, it was concluded that the results with 8192 sample window was promising compared to 16384 sample window using Levenshtein Distance measure for reference matching. A need for better approaches for jazz songs was identified during the experimentation (Klaus et al., 2019), where algorithms tracking dominant melody by a solo instrument fails to capture small duration dominant melody accompanying instruments.

Similar exploration for melodic PR using multilevel and multiple viewpoint approaches was proposed by researchers. A multilevel melodic similarity framework (Walshaw, 2015) was proposed using tune graph, an online visual tool for exploring melodic similarity. The multilevel paradigm was helpful to create a hierarchy of increasingly coarse approximations of some original representation. Evaluation of different approaches such as Edit distance, N-grams as string-based approaches, contour similarity, overall shape, ignoring ornamentation with Geometric approaches (Caraballo et al., 2022) and global similarity with features such as range, prevalent duration, intervals, mode, melodic direction, and harmonic

profile was presented (Habrard et al., 2008). The model combination and multiple viewpoint approach effectively evaluated different statistical models for melodic structures (Pearce, 2005). However, during the study of western popular music, it cannot be concluded that a strong correlation exists in the harmonic surprises and popularity of the songs, and proposed the need for behavioral studies for identifying such correlations (Miles et al., 2017).

2.5 Software tools developed by the music research community

The research community has developed various software tools for music computing applications such as analysis, synthesis, visualization and audio effects. The software and tools support different music descriptors, and feature values can be obtained with such tools. The features generated by these tools are generally low-level features, representing some statistical values extracted from the signal. Information about the different features supported by a particular tool is available through the relevant documentation. The detailed list of tools with description is available on websites like Sound and Music Computing (SMC, 2022)). Similar tools are made available by different research groups such as (ISMIR, 2022), Microsoft (MS, 2022), etc. Most of the software tools mentioned are under general public license and available on their respective web domains for free download and usage. Various audio processing plugins, typically called VAMP plugins developed by different research groups, are available for audio analysis (Vamp, 2022). Beatroot (for the beat, onset, tempo estimation), Marsyas (timbral and spectral features), Melodia (melodic pitch contour) (Melodia, 2022), and Sonic Visualizer (sonic, 2022) are some to be mentioned from the list. Audio tools developed in Matlab (Matlab, 2022) require Matlab as an essential prerequisite that is not under a general public license.

Tools like Praat (Praat, 2022) or Aubio (Aubio, 2022) allow users to adjust default parameter values, which involves choices associated with the processing of the audio data to extract appropriate features. Musical audio features are obtained by processing the audio signal with successive frames or windows. Window or frame size is one of the crucial decisions that need to be made during music processing. The window size is the length of the music file that is considered for feature extraction and processing. It is generally about 20 to 40 milliseconds for most applications (Klapuri, 2010). Features can be either localized as the frame-level (a small segment of audio) (Mohammed, 2017) or globalized for the entire audio file (Mandel et al., 2005). Global features are calculated using aggregation methods of statistical approaches such as mean, mode or median to summarize the feature values. Aggregation methods are selected depending on the type of feature to extract pertinent global

information. Classification of features helps to understand the utility and provide a basis for the selection of features.

2.6 Modelling melodic similarity

Today, the lack of appropriate content-based music information retrieval and classification systems is likely to be a motivating and driving force for feature generation and engineering to build the musical context. Melody is the heart of many music genres such as Indian music. Indian classical music based on raga is well structured by a set of rules involved. Performance improvisation and music ornamentation are peculiar characteristics associated with Indian music. The melodic ornamentation identification task requires annotation by domain experts to validate the results, which further makes the task more difficult. Indian Popular music, on the other hand, is not bound by any rules and has varied with the influence of various music genres. Most popular music composers with formal training in Indian classical music have compositions based on raga music. The popular film songs use a variety of instruments for interludes and accompaniment of the lead singers. The songs led by singers have the melody as a significant influencing factor. Interludes in the form of melodies do play an essential role in many songs.

Melodic PR in popular songs has various dimensions such as pitch, duration, loudness and ornamentation. The research community has developed various tools for pitch extraction. Although pitch extraction is relatively less complicated in monophonic music with a single melody, modelling melodic knowledge and its abstract representation is challenging. Moreover, melodic similarity based on the conventional computational approach fails to model human perception (Shen et al., 2006), (Ren et al., 2017). Therefore, modelling and using the human perception aspect for music similarity will likely reduce the semantic gap between human cognition and musical features. Music retrieval based on melody submitted or identification of songs based on similar melody needs a melodic similarity model based on human perception to achieve acceptable results. The algorithms developed should be adaptive to melody representation for different applications. Melody transforms to represent the same melody, are very common in music to describe the same melody with pitch shifting and stretched or compressed duration. The similarity measures purely based on statistical data does not model human perception. The role of melody in establishing the song's mood and associating the melodic features with emotions is further challenging. Machine learning can be applied at various stages, from feature engineering to PR, to build and test different models and algorithms.

3.   Discussion

Various techniques are available for melodic PR along with different feature approaches and software tools. Statisticians and symbolic techniques have been traditionally used and still popularly used by researchers. Neural network-based approaches are gaining ground and results are likely to improve over a period of time with better models and huge data availability. PCD and chroma-based approach based on statistical PR, and N-gram approach based on symbolic PR were observed to be dominant approaches using Euclidean distance and string matching for melodic PR. Knowledge representation for melodies has a wide variety in terms of genres, granularity requirements, multilevel and graph-based approaches, etc. The specific application determines the knowledge representation method to be used along with the algorithms associated for similarity and classification.

The music similarity perception research will enhance the current techniques to propose a more acceptable approach with a possible novel similarity measure based on the experimentation done. The current systems can be further enhanced along with the Metadata based features or keywords and content-based parameters which can be extracted and stored in the database. The users can search in the conventional way using keywords or search a song by submitting hummed query. For keyword-based queries, the proposed system works in the same way as per the currently offered methods. Listeners can modify parameters conferring to the significance of various musical parameters for the similarity of the selected songs or uploaded queries. After query submission, a matching function calculates the matching values considering the importance to different facets set by listener. Based on the matching values computed, the outcome is prepared conferring to the ranking of likely similar songs.

The suggested similarity function uses different musical facets with the parameter adjustments. The usual values will be fixed depending on the overall perception of the mainstream of music lovers. Music similarity perception varies depending on the listeners in various means based on individual choices to pleasant-sounding aspects. Many current applications use a metadata-based similarity measure for music recommendation and retrieval (Song et al., 2012). Metadata of music files in text presentation usually includes title, albums, genre, and artist. The main limitation of this method is that it can only work if metadata is available with the music. It may produce wrong results in case the data associated is incorrect. Current metadata-based systems do not support content based retrieval. Searching

of songs based on contents (Murthy et al., 2018) is the possible application of melodic similarity based on pattern recognition.

4. Conclusion

This study presents an extensive review of the current work done in melodic PR along with different approaches, their advantages and limitations, software tools used, and knowledge representations. Melody is a complex musical phenomenon involving different musical dimensions. The study reports that current solutions have limitations in terms of scalability considering the big data in music. Universal solutions applicable to all music genres are extremely unlikely. Building computational models and algorithms to train machines to perform particular tasks poses unique challenges and new opportunities. Challenges in representation and modeling melodic similarity provides research opportunities in a fascinating domain of content-based music information retrieval.

Deep neural network models for solving the PR problems are becoming increasingly popular. They can primarily learn complex non-linear input and output relationships, have almost no dependence on domain-specific knowledge, and the availability of practical models. For the effective implementation of this strategy, a vast training dataset is required. Datasets provide the foundation for machine learning algorithms. Scarcity of authentic big datasets in the public domain is one of the major issues in music computing. The involved copyright issues limit the length of the audio samples to be distributed in the public domain. Researchers are working on these issues, and deep neural network models with present variants such as transformer models will provide robust solutions in the coming years. Computational musicology is relatively a young research domain and is likely to evolve with the latest technology trends and increased computing power with GPUs. Automatic melodic pattern recognition and similarity with modeling human perception is the need of various applications such as recommendation, playlist generation, classification, and clustering. Scalable solutions with better algorithmic solutions will be the need in coming years considering the huge world music canvas.

References

[1]   Albus, John Edward, Robert H. Anderson, J. M. Brayer, R. DeMori, H-YF Feng, S. L. Horowitz, B. Moayer et al. "Syntactic pattern recognition, applications." Vol. 14. Springer Science & Business Media, 2012.

[2] Aubio audio toolbox. https://aubio.org/. [accessed 10-11-2022].

[3] Aucouturier, Jean-Julien, and Mark Sandler. "Finding repeating patterns in acoustic musical signals: Applications for audio thumbnailing." In Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio. Audio Engineering Society, 2002.

[4] Bainbridge, David, and Tim Bell. "The challenge of optical music recognition." Computers and the Humanities 35, no. 2 (2001): 95-121.

[5] Bee Suan Ong, Emilia G´omez, and Sebastian Streich. "Automatic extraction of musical structure using pitch class distribution features." In Workshop on learning the semantics of audio signals (LSAS), pages 53–65, 2006.

[6] Benetos, Emmanouil, Simon Dixon, Dimitrios Giannoulis, Holger Kirchhoff, and Anssi Klapuri. "Automatic music transcription: challenges and future directions." Journal of Intelligent Information Systems 41, no. 3 (2013): 407-434.

[7] Bishop, Christopher M., and Nasser M. Nasrabadi. "Pattern recognition and machine learning." Vol. 4, no. 4. New York: springer, 2006.

[8] Boulanger-Lewandowski, Nicolas, Yoshua Bengio, and Pascal Vincent. "Audio Chord Recognition with Recurrent Neural Networks." In ISMIR, pp. 335-340. 2013.

[9] Cambouropoulos, Emilios, Tim Crawford, and Costas S. Iliopoulos. "Pattern processing in melodic sequences: Challenges, caveats and prospects." Computers and the Humanities 35, no. 1 (2001): 9-21.

[10] Caraballo, Luis Evaristo, José Miguel Díaz-Báñez, Fabio Rodríguez, V. Sánchez-Canales, and Inmaculada Ventura. "Scaling and compressing melodies using geometric similarity measures." Applied Mathematics and Computation 426 (2022): 127130.

[11] Chaudhuri, Amit. "Finding the Raga: An Improvisation on Indian Music." Faber & Faber, 2021.

[12] Chen, Bo. "Music audio rhythm recognition based on recurrent neural network." Wireless Communications and Mobile Computing 2022 (2022).

[13] Chen, Ming-Tso, Bo-Jun Li, and Tai-Shih Chi. "Cnn based two-stage multi-resolution end-to-end model for singing melody extraction." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1005-1009. IEEE, 2019.

[14] Conklin, Darrell. "Discovery of distinctive patterns in music." Intelligent Data Analysis 14, no. 5 (2010): 547-554.

[15] Darrell Conklin. "Melody classification using patterns." In Second international workshop on machine learning and music, pages 37–41, 2009.

[16] Deutsch, Diana. "Binaural integration of melodic patterns." Perception & psychophysics 25, no. 5 (1979): 399-405.

[17] Dhariwal, Prafulla, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. "Jukebox: A generative model for music." arXiv preprint arXiv:2005.00341 (2020).

[18] Gonzalo Navarro. "A guided tour to approximate string matching". ACM computing surveys (CSUR), 33(1):31–88, 2001.

[19] Gulati, Sankalp, Joan Serra, Vignesh Ishwar, Sertan Sentürk, and Xavier Serra. "Phrase-based 285aga recognition using vector space modeling." In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 66-70. IEEE, 2016.

[20] Habrard, Amaury, José Manuel Inesta, David Rizo, and Marc Sebban. "Melody recognition with learned edit distances." In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), pp. 86-96. Springer, Berlin, Heidelberg, 2008.

[21] Ismir software tools. https://www.ismir.net/resources/software-tools/. [accessed 10-7-2022].

[22] Jagtap, Jayant, and Nilesh Bhosle. "A comprehensive survey on the reduction of the semantic gap in content-based image retrieval." International Journal of Applied Pattern Recognition 6, no. 3 (2021): 254-271.

[23] Jain, Anil K., Robert P. W. Duin, and Jianchang Mao. "Statistical pattern recognition: A review." IEEE Transactions on pattern analysis and machine intelligence 22, no. 1 (2000): 4-37.

[24] Joan Serra, Emilia G´omez, Perfecto Herrera, and Xavier Serra. "Chroma binary similarity and local alignment applied to cover song identification." IEEE Transactions on Audio, Speech, and Language Processing, 16(6):1138–1151, 2008.

[25] Jonas Langhabel, Robert Lieck, Marc Toussaint, and Martin Rohrmeier. "Feature discovery for sequential prediction of monophonic music." In ISMIR, pages 649–656, 2017.

[26] Justin Salamon. "Chroma-based predominant melody and bass line extraction from music audio signals." Unpublished master thesis, Universitat Pompeu Fabra, 2008.

[27] Khulusi, Richard, Jakob Kusnick, Christofer Meinecke, Christina Gillmann, Josef Focht, and Stefan Jänicke. "A survey on visualizations for musical data." In Computer Graphics Forum, vol. 39, no. 6, pp. 82-110. 2020.

[28] Klapuri, Anssi. "Pattern induction and matching in music signals." In International symposium on computer music modeling and retrieval, pp. 188-204. Springer, Berlin, Heidelberg, 2010.

[29] Klaus Frieler, Dogac Basaran, Frank H¨oger, H´el`ene-Camille Crayencour, Geoffroy Peeters, and Simon Dixon. "Don't hide in the frames: Note-and pattern-based evaluation of automated melody extraction algorithms." In 6th International Conference on Digital Libraries for Musicology, pages 25–32, 2019.

[30] Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering 160, no. 1 (2007): 3-24.

[31] Lee, Junghyuk, and Jong-Seok Lee. "Music popularity: Metrics, characteristics, and audio-based prediction." IEEE Transactions on Multimedia 20, no. 11 (2018): 3173-3182.

[32] Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. "A survey of deep neural network architectures and their applications." Neurocomputing 234 (2017): 11-26.

[33] Lu, Lie, Muyuan Wang, and Hong-Jiang Zhang. "Repeating pattern discovery and structure analysis from acoustic music data." In Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pp. 275-282. 2004.

[34] Lu, Wei-Tsung, and Li Su 1. "Deep learning models for melody perception: An investigation on symbolic music data." In 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1620-1625. IEEE, 2018.

[35] Lu, Wei Tsung, and Li Su 2. "Vocal Melody Extraction with Semantic Segmentation and Audio-symbolic Domain Transfer Learning." In ISMIR, pp. 521-528. 2018.

[36] Mandel, Michael I., and Daniel PW Ellis. "Song-level features and support vector machines for music classification." (2005): 594-599.

[37] Matlab audio toolbox. https://in.mathworks.com/products/audio.html. [accessed 10-11-2022].

[38] Melodia vamp plugin. https://www.upf.edu/web/mtg/melodia [accessed 10-11-2022].

[39] Microsoft computational tools for music. https://www.microsoft.com/en-us/research/project/computational-tools-for-music. [accessed 10-11-2022].

[40] Miles, Scott A., David S. Rosen, and Norberto M. Grzywacz. "A statistical analysis of the relationship between harmonic surprise and preference in popular music." Frontiers in Human Neuroscience 11 (2017): 263.

[41] Mohammed, Duraid Yehya. "Overlapped speech and music segmentation using singular spectrum analysis and random forests." University of Salford (United Kingdom), 2017.

[42] Murthy, YV Srinivasa, and Shashidhar G. Koolagudi. "Content-based music information retrieval (cb-mir) and its applications toward the music industry: A review." ACM Computing Surveys (CSUR) 51, no. 3 (2018): 1-46.

[43] Norgaard, Martin, and Ute Römer. "Patterns in music: How linguistic corpus analysis tools can be used to illuminate central aspects of jazz improvisation." Jazz Education in Research and Practice 3, no. 1 (2022): 3-26.

[44] Ong, Bee Suan, Emilia Gómez, and Sebastian Streich. "Automatic extraction of musical structure using pitch class distribution features." In Workshop on learning the semantics of audio signals (LSAS), pp. 53-65. 2006.

[45] Oramas, Sergio, Francesco Barbieri, Oriol Nieto Caballero, and Xavier Serra. "Multimodal deep learning for music genre classification." Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21. (2018).

[46] Orio, Nicola, and Antonio Rodà. "A Measure of Melodic Similarity based on a Graph Representation of the Music Structure." In ISMIR, pp. 543-548. 2009.

[47] Ozcan, Giyasettin, Cihan Isikhan, and Adil Alpkocak. "Melody extraction on MIDI music files." In Seventh IEEE International Symposium on Multimedia (ISM'05), pp. 8-pp. Ieee, 2005.

[48] Panteli, Maria, Emmanouil Benetos, and Simon Dixon. "A review of manual and computational approaches for the study of world music corpora." Journal of New Music Research 47, no. 2 (2018): 176-189.

[49] Panteli, Maria, Emmanouil Benetos, and Simon Dixon. "Learning a feature space for similarity in world music." ISMIR, 2016.

[50] Pearce, Marcus Thomas. "The construction and evaluation of statistical models of melodic structure in music perception and composition." PhD diss., City University London, 2005.

[51] Pearce, Marcus, and Martin Rohrmeier. "Musical syntax II: Empirical perspectives." In Springer handbook of systematic musicology, pp. 487-505. Springer, Berlin, Heidelberg, 2018.

[52] Pelchat, Nikki, and Craig M. Gelowitz. "Neural network music genre classification." Canadian Journal of Electrical and Computer Engineering 43, no. 3 (2020): 170-173.

[53] Praat software. https://www.fon.hum.uva.nl/praat/. [accessed 10-11-2022].

[54] Rao, Preeti, Joe Cheri Ross, Kaustuv Kanti Ganguli, Vedhas Pandit, Vignesh Ishwar, Ashwin Bellur, and Hema A. Murthy. "Classification of melodic motifs in raga music with time-series matching." Journal of New Music Research 43, no. 1 (2014): 115-131.

[55] Reddy M, Gurunath, K. Sreenivasa Rao, and Partha Pratim Das. "Melody Extraction from Polyphonic Music by Deep Learning Approaches: A Review." arXiv e-prints (2022): arXiv-2202.

[56] Ren, Iris Yuping, Hendrik Vincent Koops, Anja Volk, and Wouter Swierstra. "In search of the consensus among musical pattern discovery algorithms." In Proceedings of the 18th International Society for Music Information Retrieval Conference, pp. 671-678. ISMIR press, 2017.

[57] Rizo, David, Pedro J. Ponce De León, Carlos Pérez-Sancho, Antonio Pertusa, and José Manuel Iñesta Quereda. "A Pattern Recognition Approach for Melody Track Selection in MIDI Files." In ISMIR, pp. 61-66. 2006.

[58] Roads, Curtis, and Paul Wieneke. "Grammars as representations for music." Computer Music Journal (1979): 48-55.

[59] Rohrmeier, Martin, and Marcus Pearce. "Musical syntax I: Theoretical perspectives." In Springer handbook of systematic musicology, pp. 473-486. Springer, Berlin, Heidelberg, 2018.

[60] Sadie, Stanley, and John Tyrrell. "Dictionary of music and musicians." New York: Oxford University Press. Yónatan Sánchez, 2001.

[61] Salamon, Justin, Emilia Gómez, Daniel PW Ellis, and Gaël Richard. "Melody extraction from polyphonic music signals: Approaches, applications, and challenges." IEEE Signal Processing Magazine 31, no. 2 (2014): 118-134.

[62] Samuel Kim and Shrikanth Narayanan. "Dynamic chroma feature vectors with applications to cover song identification." In 2008 IEEE 10th Workshop on Multimedia Signal Processing, pages 984–987, 2008.

[63] Shen, Jialie, John Shepherd, and Anne HH Ngu. "Towards effective content-based music retrieval with multiple acoustic feature combination." IEEE Transactions on Multimedia 8, no. 6 (2006): 1179-1189.

[64] Shmulevich, Ilya, Olli Yli-Harja, Edward Coyle, Dirk-Jan Povel, and Kjell Lemström. "Perceptual issues in music pattern recognition: Complexity of rhythm and key finding." Computers and the Humanities 35, no. 1 (2001): 23-35.

[65] Song, Yading, Simon Dixon, and Marcus Pearce. "A survey of music recommendation systems and future perspectives." In 9th international symposium on computer music modeling and retrieval, vol. 4, pp. 395-410. 2012.

[66] Sonic visualizer. https://www.sonicvisualiser.org/. [accessed 10-11-2022].

[67] Sound and music computing software tools. https://smcnetwork.org/software.html. [accessed 10-11-2022].

[68] Subramanian, Aswin Shanmugam, Chao Weng, Shinji Watanabe, Meng Yu, and Dong Yu. "Deep learning based multi-source localization with source splitting and its effectiveness in multi-talker speech recognition." Computer Speech & Language 75 (2022): 101360.

[69] Thickstun, John, Zaid Harchaoui, and Sham Kakade. "Learning features of music from scratch." arXiv preprint arXiv:1611.09827 (2016).

[70] Tuomas Eerola, Topi J¨a¨arvinen, Jukka Louhivuori, and Petri Toiviainen. "Statistical features and perceived similarity of folk melodies." Music Perception, 18(3):275–296, 2001

[71] Ullrich, Karen, Jan Schlüter, and Thomas Grill. "Boundary Detection in Music Structure Analysis using Convolutional Neural Networks." In ISMIR, 2014.

[72] Vamp plugins for software tools. https://www.vamp-plugins.org/download.html. [accessed 10-11-2022].

[73] Vempala, Naresh N., and Frank A. Russo. "Predicting emotion from music audio features using neural networks." In Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR), pp. 336-343. London, UK: Lecture Notes in Computer Science, 2012.

[74] Walshaw, Chris. "Multilevel melodic matching." In 5th Intl. Workshop on Folk Music Analysis, pp. 130-137. 2015.

[75] Wang, Xiaolu, and Yao Chen. "Music teaching platform based on FPGA and neural network." Microprocessors and Microsystems 80 (2021): 103337.

[76] Xiang Cao. "Automatic accompaniment of vocal melodies in the context of popular music." PhD thesis, Georgia Institute of Technology, 2009.

Author's biography

Makarand Velankar was born in 1969 and completed basic graduation in computer engineering from Walchand college of Engineering, Sangli in 1990. After working in the industry for 11 years as development engineer, he joined MKSSS's Cummins College of engineering, Pune in 2001 and is presently working as Assistant Professor in Information Technology Department. He completed PhD from Savitribai Phule Pune University (SPPU) in the domain of Computational musicology. His research interests include music analysis, machine learning, artificial intelligence, Information retrieval, algorithms, soft computing etc.

Parag Kulkarni is one of the world's leading authorities on Business Strategy, Knowledge Innovation, Machine Learning, Systemic Learning and Building Innovative Knowledge Corporations in the knowledge economy. He is consultant on Innovation and Strategies for start-ups and SMEs and contributed to make many start-ups successful. He holds PhD from IIT Kharagpur, Management education from IIM Kolkata. Parag is Founder, Chief Scientist and CEO of the iknowlation Research Labs Pvt Ltd: an innovation, strategy and business consulting and product development organization. His core work includes innovation and knowledge strategies for start-ups, knowledge strategies for organizations, building innovative products, breaking away from competitive landscape. His works are on systemic innovation and learning, which are published in many reputed journals.